doi: 10.1046/j.1529-8817.2004.00096.x

Mitochondrial DNA Heterogeneity in Tunisian Berbers

K. Fadhlaoui-Zid[1,*], S. Plaza[2,*], F. Calafell[2], M. Ben Amor[1], D. Comas[2,**] and A. Bennamar El gaaied[1]

[1] *Laboratoire de Génétique Moléculaire, Immunologie et Biotechnologie, Faculté des Sciences de Tunis, Université Tunis El Manar, 2092 Tunis, Tunisia*

[2] *Unitat de Biologia Evolutiva. Universitat Pompeu Fabra. Doctor Aiguader 80, 08003 Barcelona, Spain*

Summary

Berbers live in groups scattered across North Africa whose origins and genetic relationships with their neighbours are not well established. The first hypervariable segment of the mitochondrial DNA (mtDNA) control region was sequenced in a total of 155 individuals from three Tunisian Berber groups and compared to other North Africans. The mtDNA lineages found belong to a common set of mtDNA haplogroups already described in North Africa. Besides the autochthonous North African U6 haplogroup, a group of L3 lineages characterized by the transition at position 16041 seems to be restricted to North Africans, suggesting that an expansion of this group of lineages took place around 10500 years ago in North Africa, and spread to neighbouring populations. Principal components and the coordinate analyses show that some Berber groups (the Tuareg, the Mozabite, and the Chenini-Douiret) are outliers within the North African genetic landscape. This outlier position is consistent with an isolation process followed by genetic drift in haplotype frequencies, and with the high heterogeneity displayed by Berbers compared to Arab samples as shown in the AMOVA. Despite this Berber heterogeneity, no significant differences were found between Berber and Arab samples, suggesting that the Arabization was mainly a cultural process rather than a demographic replacement.

Introduction

Berbers inhabit scattered places in North Africa, from the Moroccan western coast to the oasis Siwa in Egypt, and from Tunisia in the north to the oases in the mid-Sahara. The origin of the Berber people is not clearly established. According to the archaeological record, North Africa has been peopled since Upper Palaeolithic times (Newman, 1995). The first well-defined Palaeolithic technology, the Aterian, dates back around 40,000 years ago, and is followed by the Iberomaurisian (∼22,000 years ago; Feremback, 1985; Close & Wendorf, 1990). The archaeological record then reveals a Mesolithic culture, the Capsian (Brett & Fentress, 1996), which gave way to the Neolithic transition to agriculture that occurred around 9,500-7,000BC, spreading from the Near East to Egypt (Dupanloup, 1993). Berbers may be the descendants of Mesolithic Capsian populations, and/or of the later Neolithic people who came from the Middle East via Egypt and who possibly introduced the Afro-Asiatic languages to North Africa (Renfrew, 1991). Since then, the North African coast has known several invasions: Phoenicians, Romans, Vandals, Byzantines, Arabs, Ottomans, Spanish and French have occupied the territory, although their demographic impact is not well established.

In Tunisia, the first well-known post-Neolithic invasion was that of the Phoenicians coming from the East Mediterranean sea coast around 1,100BC. Nonetheless, their number was estimated at the end of their kingdom to be 100,000 compared to 500,000 Berbers living in Tunisia (Julien, 1961). The long dominations by the Romans, Vandals and Byzantines had an even lesser demographic impact. The Arab conquest in Tunisia started in the 7th century and was followed by a massive Bedouin

*Both authors contributed equally to the present work

**Correspondence: David Comas, Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Doctor Aiguader 80, 08003 Barcelona, Spain. Tel: +34 93 542 28 02; Fax: +34 93 542 28 44. E-mail: david.comas@upf.edu

immigration during the 11th century. During this invasion, Berbers were restricted to Numidia (the Centre and South of Tunisia). The Arab expansion largely submerged the original Berber language and customs, except for the tribes that were forced back to the mountains and certain villages located in Southern Tunisia. Between the sixteenth and twentieth centuries Tunisia was ruled by Turks, Spaniards and French. Other immigrants came from the South (Sub-Saharan African slaves from Sudan). All of these populations probably contributed to the present Tunisian gene pool.

Nowadays, the only criterion used to distinguish the Berbers from the rest of the Tunisian population is the language called Chelha. Berbers who speak Chelha in Tunisia are 1% of the global population, and are localized in four villages in the South of Tunisia (Sened, Matmata, Chenini and Douiret) and in the island of Jerba.

Mitochondrial DNA (mtDNA) is a powerful tool in reconstructing population history, because a fine-grained phylogeography has been defined for it. This is particularly relevant in Northern Africa, which is at the crossroads of Europe, the Middle East, and Sub-Saharan Africa, regions for which mtDNA phylogeography is known in detail (Macaulay *et al.* 1999; Richards *et al.* 2000; Salas *et al.* 2002).

Several genetic studies have been performed in North African populations, although very few in Tunisian Berbers. The compilation of classical genetic markers in North Africa (Bosch *et al.* 1997) showed a clear genetic differentiation between East and West, attributed to human expansions from the Middle East, such as Palaeolithic and/or Neolithic demographic expansions. Analyses of autosomal STRs (Bosch *et al.* 2000) and *Alu* insertion polymorphisms (Comas *et al.* 2000) showed some Sub-Saharan genetic flow into NW African populations. High-resolution analysis of Y-chromosome biallelic and STR markers (Bosch *et al.* 2001) has revealed a clear North African genetic differentiation compared to Europe, due to a major independent Upper Palaeolithic contribution in both areas, followed by gene flow from the Near East during the Neolithic, and small bidirectional gene flow across the Mediterranean. Finally, mtDNA in North Africa has been analysed mainly in the Western coast (Rando *et al.* 1998; Brakez *et al.* 2001; Plaza *et al.* 2003) and Egypt (Krings *et al.* 1999), and has been studied only in a non-Berber population in Tunisia (Plaza *et al.* 2003). Previous mtDNA analyses have suggested that modern Berbers are the descendants of the earlier groups living in North Africa in Palaeolithic times (Rando *et al.* 1998; Macaulay *et al.* 1999). The analysis of GM and KM haplotypes in Tunisian Berbers (Chaabani *et al.* 1984) suggested some heterogeneity within Berbers; thus mtDNA analysis will contribute significantly to our knowledge of the genetic pool of Tunisian Berbers.

In the present study we have analyzed mtDNA HVS-I sequences in three Berber isolates from Southern Tunisia, with the aim of evaluating the possible heterogeneity of these Berbers in relation to genetic drift, comparing them to several Arabic and Berber-speaking populations from North Africa, and determining the genetic contributions of surrounding populations to the Tunisian Berbers, to trace their population history.

Materials and Methods

Samples

A total of 155 Berbers from Tunisia were analysed for the hypervariable segment I (HVS-I) of the mtDNA non-coding region. Blood samples were collected from four villages (Figure 1): 53 from Sened (also known as Sundia), 49 from Matmata (or Matmatia), and 53 from Chenini (Chenenaouia) and Douiret (Douiria). Chenini and Douiret are two neighbouring villages 20 km from each other; samples from these two villages were pooled and treated as a single population. Blood samples were collected according to geographic and linguistic criteria: donors were Berber Chelha speakers born in one of the four villages mentioned above. DNA extraction was performed using a standard phenol-chloroform method.

MtDNA Amplification and Sequencing

The HVS-I was amplified using the primers L15996 and H16401 as described in Vigilant *et al.* (1989). PCR products were purified with the QiAEX II KIT (Qiagen). The sequencing reaction was performed using the Big Dye Terminator (version 3.0) Cycle Sequencing Kit, with AmpliTaq® DNA Polymerase (Applied Biosystems). Sequences were run

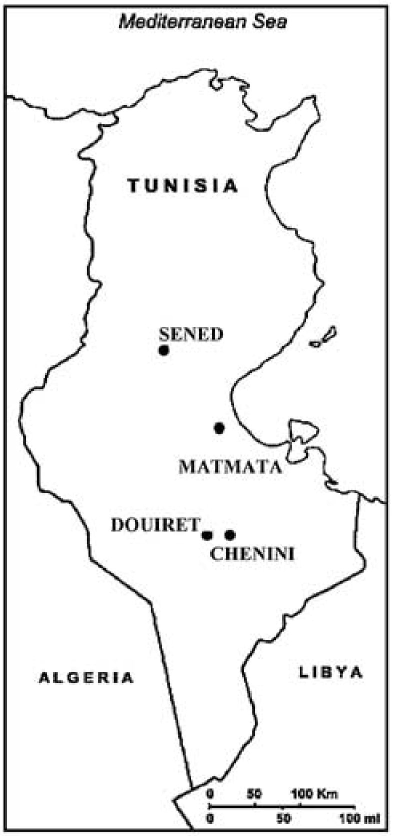


Figure 1 Geographical location of the localities sampled.

in an automatic Sequencer ABI377 (Applied Biosystems). Sequence analyses were performed from positions 16024 to 16391 according to the Cambridge Reference Sequence (CRS; Anderson *et al.* 1981; Andrews *et al.* 1999), and are available at http://www.upf.es/cexs/recerca/bioevo/index.htm

In addition to the control region sequence, four single nucleotide polymorphisms (SNPs) in the mtDNA coding region (positions 10400, 10873, 12308, and 12705) were determined in those individuals that were difficult to classify unambiguously into haplogroups based only in the information provided by the HVS-I. A single-base primer elongation method was used to genotype these four positions, and primers and PCR conditions are described elsewhere (Comas *et al.* 2004). Information yielded by each of the four SNPs allowed us to assign sequences to one of four major haplogroups: 10400T defines the major M haplogroup, 10873C defines the major L haplogroup, 12308G defines the major U haplogroup (including the K haplogroup), and 12705C defines the major R haplogroup, which includes a large set of haplogroups (H, V, J, T, U, B and F).

Phylogenetic Analyses

DNA sequences were aligned using the CLUSTAL V program. Sequences were classified into haplogroups according to the nomenclature of Richards *et al.* (2000) and Salas *et al.* (2002). Genetic diversity measures were calculated (from position 16024 to 16383) with the Arlequin package 2.0 (Excoffier *et al.* 1992) and compared to those in a set of North African populations (Table 1). Analyses of molecular variance (AMOVA) were also performed with Arlequin 2.0. Genetic distances between populations using the HVS-I were calculated by intermatch-mismatch pairwise differences according to the equation $D = d_{ij} - (d_{ii} + d_{jj})/2$ (Nei, 1987), where d_{ij} is the mean pairwise differences between populations i and j, and d_{ii} and d_{jj} are the mean pairwise differences within populations i and j respectively. The distance standard errors were computed by resampling nucleotide positions with 1,000 bootstrap iterations (Efron, 1982). The distance matrix was also employed to build a multidimensional scaling using the STATISTICA 6.0 package. Principal component and correspondence analyses were obtained from haplotype frequencies using the SPSS package.

The network relating HVSI sequences within some of the haplogroups described was constructed by using a reduced-median algorithm (Bandelt *et al.* 1995) as implemented in the Network 3.0 program. The dating method employed (Saillard *et al.* 2000) is based on the average number of mutations accumulated from an ancestral sequence as a linear function of time and mutation rate. This method was also performed with the Network 3.0 program.

Results

Haplogroup assignation was performed comparing HVS-I sequences with the data set of Richards *et al.* (2000) and Salas *et al.* (2002), and their classification was used. All Chenini-Douiret sequences were classified unambiguously using the information yielded by the HVS-I variable positions. One sequence from Sened

Table 1 Genetic diversity parameters in North African samples

Populations	Code	Ref.	n	H (SE)	π
Matmata Berbers	TB M	1	49	0.964 ± 0.021	5.05
Sened Berbers	TB S	1	53	0.975 ± 0.011	7.53
Chenini-Douiret Berbers	TB CD	1	53	0.939 ± 0.017	6.82
Tunisian Arabs	Tun	2	47	0.990 ± 0.009	6.15
Mozabites	Moz	3	85	0.942 ± 0.010	4.73
Algerian Arabs	Alg	2	47	0.956 ± 0.014	5.72
Center Moroccan Berber	CB1	2, 4	64	0.968 ± 0.013	4.51
Center Moroccan Berber	CB2	8	60	0.984 ± 0.009	6.00
South Moroccan Berbers	SB	5	50	0.961 ± 0.018	4.60
Moroccan Arabs	MA	2, 4	50	0.993 ± 0.006	7.04
Mauritanians	Mau	4	30	0.975 ± 0.017	6.09
Egyptians	Eg	6	68	0.992 ± 0.005	7.06
Saharawi	Sah	2, 4	81	0.982 ± 0.006	5.44
Tuareg	Tg	7	26	0.985 ± 0.014	7.10

H: sequence diversity
π: average number of pairwise differences
References: 1: Present study; 2: Plaza *et al.* (2003); 3: Côrte-Real *et al.* (1996), 4 : Rando *et al.* (1998), 5: Brakez *et al.* (2001), 6 : Krings *et al.* (1999), 7: Watson *et al.* (1996), 8 : Thomas *et al.* (2002)

and nine from Matmata that could not be unambiguously assigned into haplogroups based only on the data from the control region were typed for positions 10400, 10873, 12308, and 12705 in the coding region. From these, the sequence from Matmata with the following mutational pattern 16232-16293-16356 could not be typed for these four positions due to lack of DNA, and was classified as "other." Haplogroup frequencies are reported in Table 2.

Sequence Gene Pool in Tunisian Berbers

The Tunisian Berber mtDNA gene pool is constituted by sequences belonging to several major haplogroups: the sub-Saharan African L group, the east African M1 haplogroup, and the West Eurasian major haplogroups N and R.

Sub-Saharan sequences, represented by haplogroups L1, L2, and L3, are found in the three Tunisian Berber populations at high frequencies: 26.6% in the Berbers from Sened, 24.3% in the Berbers from Matmata, and 13.3% in the Berbers from Chenini-Douiret, showing a large sub-Saharan gene flow among these three Berber populations. Most of the sub-Saharan sequences found were not described in the data sets of Salas *et al.* (2002), although most of them differ from an already described sequence by one or two mutation steps. Noteworthy is the presence of eight sequences, out of the 17 L3 sequences found in Tunisian Berbers, that harbour a transition at nucleotide position 16041. This transition has been found in other Berber and Arab samples from North Africa, and its presence is negligible in other populations (two South-Western Europeans, three West Africans and one East African), suggesting a North African origin for this group of sequences, although a sub-Saharan origin cannot be rejected. The network of sequences bearing the 16041 transition (Figure 2) shows a clear star-like phylogeny and a recent origin for this group, at around 10500 years (SE 3500 years).

The M1 haplogroup, to which an East-African origin is attributed (Quintana-Murci *et al.* 1999), is represented by a single sequence in six Sened Berbers (10.9%), in one individual from Matmata, and it is absent in Berbers from Chenini-Douiret. This sequence was found among Arabs from Tunisia but not in Algerians (Plaza *et al.* 2003). This sequence was also found in other populations (Quintana-Murci *et al.* 1999; Richards *et al.* 2002) with a transversion at position 16183, which may be the result of the hypervariable length polymorphism in the poly-C tract of the control region (Bendall & Sykes, 1995). The presence of a single M1 sequence at high frequency in Tunisian Berbers could be explained by gene flow from East Africa, followed by genetic drift.

Table 2 MtDNA haplogroup frequencies in Tunisia

Haplogroups	TB CD N = 53	TB S N = 53	TB M N = 49	Tunisian Arabs[a] N = 47
L1b*	–	–	2.0	–
L1b1	–	5.7	–	–
L1c1	3.8	–	–	–
L2*	–	1.9	–	2.1
L2a	1.9	1.9	2.0	4.3
L2a1	–	3.8	–	4.3
L2a1a	1.9	3.8	2.0	–
L2b1	–	–	–	2.1
L3*	–	1.9	16.3	2.1
L3b	3.8	3.8	–	–
L3b1	–	1.9	–	–
L3d	–	–	2.0	2.1
L3e1	–	–	–	2.1
L3e2	–	1.9	–	4.3
L3f	1.9	–	–	4.3
M1	–	11.3	2.0	4.3
N1b	1.9	–	8.2	–
I2	–	7.5	–	–
W	–	–	–	2.1
X	–	–	–	2.1
pre-HV	–	5.7	–	–
HV	15.1	–	–	6.4
H*	13.2	24.5	26.5	23.4
V	–	–	16.3	–
J*	3.8	3.8	2.0	–
J1	–	–	2.0	2.1
J2	–	–	–	2.1
T*	1.9	1.9	–	–
T1	30.2	1.9	4.1	2.1
T2	–	–	–	2.1
T3	–	–	–	2.1
K*	5.7	–	4.1	4.3
K2	9.4	–	–	2.1
U*	–	1.9	–	–
U1a	–	3.8	–	4.3
U3	5.7	–	2.0	–
U5*	–	1.9	–	–
U5a1a	–	1.9	–	–
U5b	–	–	–	2.1
U6a*	–	1.9	2.0	4.3
U6a1	–	5.7	–	–
U7a	–	–	4.1	–
Other	–	–	2.0	6.4

[a]Data from Plaza *et al.* 2003

The U6 haplogroup originated in N Africa ∼40,000 years ago (Macaulay *et al.* 1999) and is found in Moroccan Berbers at 6–8%, reaching 28% in the isolated Mozabites from Algeria. In Tunisian Berbers it is found at 7.6% in Sened, 2.0% in Matmata and was absent in Chenini-Douiret; it has been found at 4.2% in Tunisian Arabs (Plaza *et al.* 2003). Thus, U6 frequencies in Tunisian Berbers are relatively low and may mark an eastward decline in the frequencies of this haplogroup. All U6 lineages found in Tunisian Berbers belong to the U6a (characterised by 16278T), and U6a1 (characterised by 16189C and by 16278T) subgroups.

Haplogroup U/K, excluding U6 sequences, was found in 9.5% of the Sened sequences, 10.2% in the Matmata and 20.8% in Chenini-Douiret. Haplogroup U is represented by U1a, U3, U5, U5a1a, U7a, and K sequences. The high frequency of this haplogroup in the Chenini-Douiret sample is due to the presence of eight sequences belonging to haplogroup K. Three of them belong to the K root type, and the other five to the haplogroup K2, which is represented by two haplotypes. Haplogroup U7a reached a frequency of 4.1% in the Berbers from Matmata. This haplogroup is found in Middle Eastern populations such as Iraqis, Palestinians, Armenians, Druze and Kurds; and some Southern and Eastern Europeans (Richards *et al.* 2000), but none of the already described sequences matched with those found in the Matmata Berbers. U7a sequences may have been integrated into Matmata Berbers after a wave of migration from the Middle East, as they differ by only one step mutation from the U7a Middle Eastern sequences, whereas three mutation steps separate them from the European U7a sequences. Haplogroup U3 is mainly found in populations from the Middle East (Richards *et al.* 2000), and was also reported in Chenini-Douiret Berbers at 5.7%, and in Matmata Berbers at 2.0%. But no match was found between Tunisian Berbers and Middle Eastern U3 sequences, nor between U3 sequences among Tunisian Berbers.

H is the most frequent haplogroup in most West Eurasian (Richards *et al.* 2000) and North African (Plaza *et al.* 2003) populations, as well as in Tunisian Berbers, where H sequences (and HV sequences, since sometimes they cannot be directly distinguished neither by HVS-I nor by the four coding positions typed in the present study, and the typing of position 7025 would be recommended) are found at a high frequency: 28.3% in the Berbers from Chenini-Douiret, 24.5% in the Berbers from Sened, and 26.5% in the Berbers from Matmata. Haplogroup V, which is largely distributed in Western Mediterranean populations (6% in NW Africa,

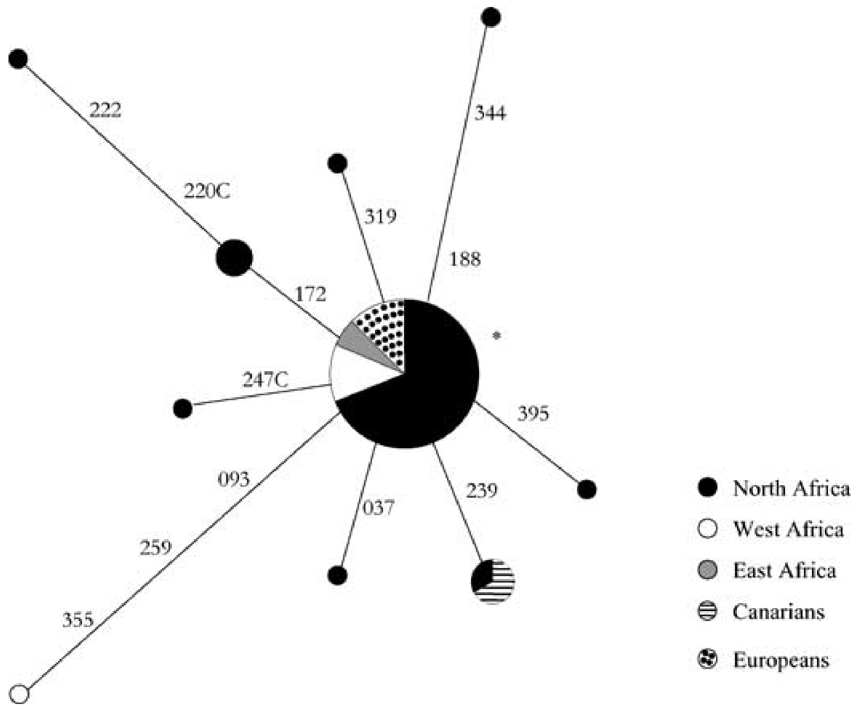


Figure 2 Phylogenetic network of a section of haplogroup L3 bearing a transition in position 16041. The size of the circles is proportional to the number of sequences. Mutated sites (minus 16000) are indicated along the lines. Transversions are indicated by the nucleotide after the number. The sequence marked with an asterisk differs from the Cambridge Reference Sequence by transitions at positions 16041 and 16223.

5% in Iberian Peninsula, and 3,2% in Italy; Plaza *et al.* 2003), is found only in the Matmatia with a high frequency of 16.3%. Haplogroup V is absent in the two other Tunisian Berber populations, as is also the case for Tunisian and Algerian Arabs. This value is comparable with that found in the Saharawi (17.9%) and higher than those found in the Basques (10.4%) and other Western European samples (Plaza *et al.* 2003). 8.2% of the V sequences of the Matmata Berbers bear only the variant 16298C, and the rest of the sequences bear one additional variant and did not match with previously described V sequences (Richards *et al.* 2000). It is also relevant to note the high frequency of haplogroup HV1 (15.1%) in Berbers from Chenini-Douiret, represented by a unique sequence.

Haplogroup J is found at very low frequencies in the three samples, and except for one Matmata sequence that can be classified into the J1 group, the rest of sequences remain in the J root type. All the T sequences found in the present sample set, except one individual from Sened and one from Chenini-Douiret who bear a T root type, belong to the T1 subgroup. The distribution of haplogroup T in the three Tunisian Berber samples is very different. The frequency of T sequences is very low in Matmata (4.1%) and Sened Berbers (3.8%), whereas haplogroup T is represented in the Chenini-Douiret sample by 17 individuals (32.1%) carrying six different haplotypes. Of those, the most frequent was found in nine individuals and contains a deletion at position 16193, which was also observed in a single one-step derivative. This sequence type, without the deletion, was shared with Moroccan Berbers, Mozabites and Egyptians, and is also frequent in Europe and the Middle East.

Other haplogroups that are mainly found in the Middle East are also present in Tunisian Berbers: 8.2% of Matmata Berbers and one individual from Chenini-Douiret belong to N1b haplogroup, and 7.5% of Sened Berbers displayed I2 sequences.

Population Structure

An analysis of the molecular variance (AMOVA) was performed with the sample data set described in

Table 3 Analyses of Molecular Variance (AMOVA) in North African samples

Groups	Among groups	Among populations within groups	Within populations
All populations		4.14**	95.86**
Arabs		1.54**	98.46**
Berbers		6.24**	93.76**
Moroccan Berbers		2.47**	97.53**
Tunisian Berbers		6.67**	93.33**
Arabs vs Berbers	−0.21 ns	4.25**	95.95**
Arabs vs Berbers[a]	0.04 ns	3.10**	96.86**
Moroccan Berbers vs Algerian (Moz) vs Tunisian Berbers	0.60 ns	4.98**	94.42**
Moroccan Berbers vs Tunisian Berbers	−0.16 ns	4.79**	95.37**
Morocco vs Algeria-Tunisia	−0.34 ns	4.36**	95.98**

** ($p < 0.01$); * ($p < 0.05$); ns : not significant
[a]Mozabites and Tuareg excluded from the Berber group

Table 1. When all populations were considered as a single group, 4.14% of the variance was attributed to differences among populations, and 95.86% of the genetic variance was found within populations (Table 3). These populations were more diverse than sub-Saharan and West Mediterranean populations, where 98.8% (Salas *et al*. 2002) and 97.4% (Plaza *et al*. 2003) of the variance was found within the respective population sets. However, this could be due to the sampling of one or a few outlier populations that are not diluted in a high number of populations, such those used in the sub-Saharan and West Mediterranean analyses.

Several population groups based on cultural and geographic criteria were performed in order to detect some population structure. Populations were first clustered according to linguistic and cultural criteria (Arab vs. Berber), and AMOVA was performed within each group. Berber populations were more diverse from each other than Arabs were: 6.24% of the genetic variance was due to differences among Berber samples whereas 1.54% was attributed to differences among North African Arabs. When these two groups were compared, genetic variation between Arabs and Berbers was non-significantly different from 0, and 4.25% of the genetic variance was due to differences within Arabs and Berbers. When the analysis was also performed removing Mozabites and Tuaregs, known isolate Berbers where drift has acted strongly, the variation among groups remains non-significant, whereas the variation within Arabic and Berber groups decreases to 3.10% ($p < 0.01$). This result shows that, even when known outliers are removed, a certain degree of heterogeneity remains within groups, whereas no genetic differences are revealed between Arabs and Berbers.

We next classified Berber populations according to their geographic origin (Morocco, Algeria, and Tunisia), and no significant differences were found among groups. Nonetheless, the variance among populations from the same country was significantly high (4.98%, $p < 0.01$), reflecting heterogeneity among the Berber populations from the same geographical area. Since Mozabites are known outliers that could contribute disproportionately to interpopulation variance, they were removed from the analysis (Table 3). The apportionment of the variance due to differences within groups dropped to 4.79%, but was still highly significant ($p < 0.01$).

In order to detect a possible east-west differentiation in Northwest African samples suggested previously (Plaza *et al*. 2003), two geographical groups (Morocco versus Algeria-Tunisia) were formed, pooling Arabs and Berbers (Table 3). 4.36% ($p < 0.01$) of the genetic variance was attributed to differences within groups, whereas no significant differences were found between groups.

In summary, AMOVA showed that Tunisian Berbers are highly heterogeneous populations.

Tunisian Berbers within the North African Genetic Landscape

Measures of genetic diversity are reported in Table 1. Berbers from Chenini-Douiret present the lowest sequence diversity value of the geographical region considered, even lower than the diversity presented by the Mozabites, although its confidence interval overlaps widely with all others. However, the mean pairwise difference in Chenini-Douiret is not lower than in other populations. This result points to a micro geographic differentiation among the Tunisian Berbers due to genetic drift, since some of the sequences (coming from a variety of haplogroups, hence the relatively high average pairwise difference) have high frequencies in the sample. When compared with the rest of the North African samples, the two other Tunisian Berber populations displayed haplotype diversity values that are in the range of the observed values.

In order to establish the genetic relationship between the Tunisian Berbers and the other of North African populations, the genetic distance matrix between populations based on individual sequences was represented as a bidimensional plot by means of multidimensional scaling (MDS; Figure 3). The MDS plot isolates the Tuareg, the Mozabites and the Berbers from Chenini-Douiret from the rest of the populations, placing each of them at one extreme of the plot. The Berbers from Matmata are close to the South and Central Moroccan Berbers, whereas the Berbers from Sened are adjacent to Arab populations. A principal co-ordinate analysis, based on the same genetic distance matrix, yielded very similar results (data not shown).

On the basis of the frequency of haplogroups, a correspondence analysis was performed (Figure 4). The Berbers from Chenini-Douiret are associated in the plot with haplogroups T, and HV, the Tuareg with L2; and Mozabites with U6 and V, each at one extreme of the plot, whereas the rest of the populations are situated in the centre of the plot. A principal component analysis produced similar results (not shown). Thus, it seems that the heterogeneity detected by AMOVA is mostly contributed by the Chenini-Douiret sample.

Discussion

The mtDNA haplogroup composition of Tunisian Berbers reveals a similar picture to that seen in other Northern African populations (Rando *et al.* 1999; Plaza

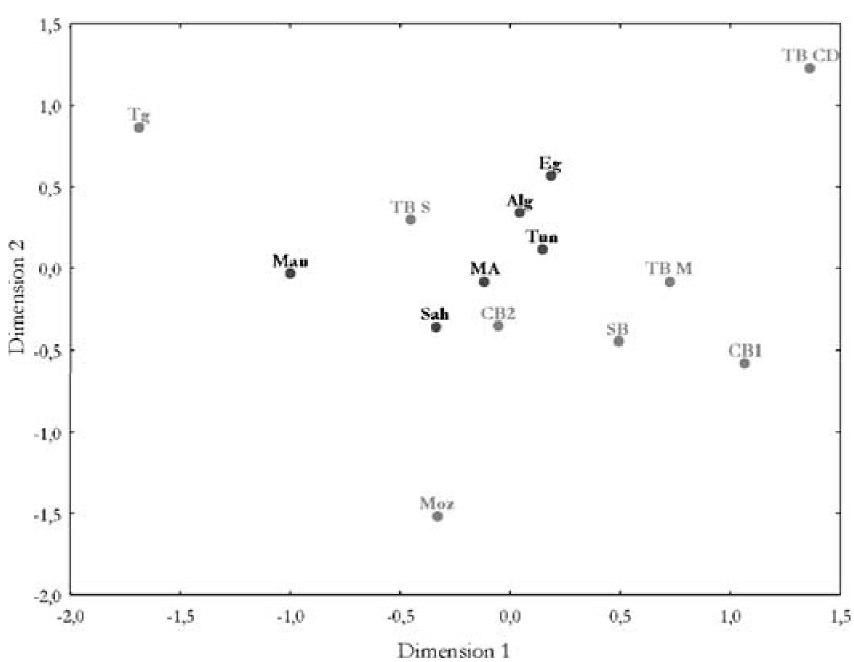


Figure 3 Multidimensional scaling (MDS) analysis based on the genetic distance matrix of North African samples. Abbreviations as in Table 1.

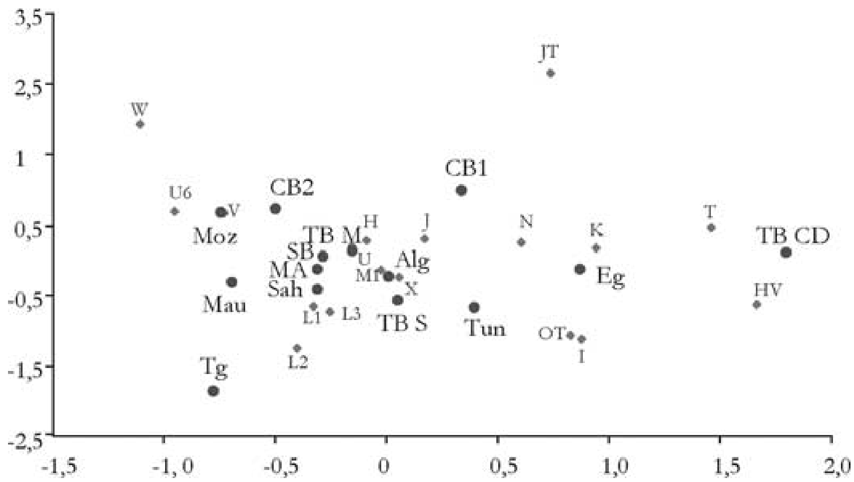


Figure 4 Plot of the analysis of correspondence based on the haplogroup frequencies of North African samples. Circles represent population samples and squares represent haplogroups. Abbreviations as in Table 1.

et al. 2003), with ~57% of sequences found in haplogroups of broad West Eurasian distribution, ~26% of sub-Saharan origin, ~14% of recent Middle Eastern origin, and ~3% locally originated in N Africa (that is, U6, although the L3 lineage with a transition at 16041 should probably be added to this category).

Few L sequences found in Tunisian Berbers were described in sub-Saharan African samples (Salas *et al.* 2002), but they differ from one or two positions from previously described lineages, providing some evidence of sub-Saharan admixture. Nevertheless, the group of sequences belonging to the L3 group with the substitution at position 16041 was also found in other North African samples, such as Moroccan Arabs, South Moroccan Berbers, and Algerians, but no match was found with sub-Saharan populations. This result points to a recent origin for this group of sequences, estimated at around 10500 years, rather than to an ancient sub-Saharan gene flow. The presence of these L3 lineages and the U6 haplogroup suggests that human populations in North Africa have experienced several population expansion processes after the occupation of the region by anatomically modern humans, and the extant populations are not only the result of external migrations from the Middle East, sub-Saharan Africa and Southern Europe into the area.

The high frequency of the East African haplogroup M1 in Sened is similar to that in Somalians (11%; Watson *et al.* 1997), Nubians (10%; Krings *et al.* 1999), and Ethiopians (10%; Thomas *et al.* 2002), nonetheless this haplogroup is represented by a single haplotype. The same pattern is shown by other Berber samples, with single haplotypes reaching high frequencies, such as two sequences belonging to the HV1 and T1 haplogroups in Chenini-Douiret, or two U6a1 sequences in the Mozabites. This result suggests that the introduction of these sequences was followed by drift, probably due to isolation of Berber populations, which led to different haplogroup frequencies and yielded a high degree of heterogeneity among Berber groups.

The three Tunisian Berber populations analysed for the mtDNA control region are characterized by high genetic heterogeneity, despite their geographic proximity and their common culture and history. Although they are separated by ~100 Km, they reveal similar haplogroups but at different frequencies. Tunisian Berbers are more heterogeneous than Moroccan Berbers, since an AMOVA within each group shows an apportionment of the variance in Moroccan Berbers of 2.47% ($p < 0.01$) compared to 6.67% ($p < 0.01$) in Tunisian Berbers. Nonetheless, when Chenini-Douiret Berbers are excluded from the analysis, similar values to those of Moroccan Berbers are found in Tunisian samples (3.97%, $p < 0.01$; data not shown). Therefore, the analyses performed place the Berbers from Chenini-Douiret as an outlier sample within Berbers, similar to what is observed in Mozabites. Nonetheless, the particularity of these Berber samples is not the result of the presence

of unusual haplogroups and/or sequences in the region (Plaza *et al*. 2003). Therefore, the outlier position of these Berber groups may be the result of a relatively recent isolation process that has caused sequence frequencies to drift to unusual values when compared to the rest of samples within the geographical area.

From the three Tunisian Berber groups, the Berbers from Chenini-Douiret seem to be the most genetically isolated with a high West Eurasian component in their mtDNA gene pool. This result is in accordance with studies based on the analysis of the polymorphic Gm system (Chaabani & Cox, 1988; Helal *et al*. 1988) of Tunisian populations, which also have a very defined phylogeography. Sub-saharan African Gm haplotypes were present at a low frequency (0.07) among the Berbers from Chenini-Douiret, whereas the West Eurasian haplotypes were much more frequent (0.62). The isolation of this Berber group may explain their variant haplogroup frequencies, low sequence diversity, and lower sub-Saharan African contribution.

The cultural differentiation present in North Africa between Berber and Arab samples seems not to reflect genetic differences between both groups, as shown in the AMOVA analyses, and the MDS and PC analyses. If Arabs in Northern Africa were mostly descendants of Middle Eastern Arabs, the frequencies of haplogroups such as N, U1, U3, U7, and HV that are much more prevalent in the Middle East than elsewhere should be larger in N. African Arabs than in Berbers. However, the opposite is observed: these haplogroups add up to 5% in N. African Arabs but to 10% in Berbers. Drift in some of the more isolated Berber populations could explain this observation. The lack of differentiation between North African Arabs and Berbers has also been observed using other genetic markers such as classical markers (Bosch *et al*. 1997); autosomal STRs (Bosch *et al*. 2000), *Alu* insertion polymorphisms (Comas *et al*. 2000); and Y-chromosome lineages (Bosch *et al*. 2001). This pattern suggests that the Arabization of the area was mainly a cultural process, rather than a demographic replacement of the Berber populations that inhabited the region where the Arabic expansion took place.

The present data has failed to confirm an east-west differentiation of North African populations as previously suggested using mtDNA sequences (Plaza *et al*. 2003) or other genetic markers (Bosch *et al*. 1997). The present mtDNA data show a more patchy genetic landscape in North Africa, with some Berber samples acting as outliers in the general North African landscape. The lack of mtDNA data for large geographic regions like the Kabylie (Algeria) and Libya, and the large number of isolated Berber samples considered in the present analysis may decrease the power to find the longitudinal differentiation previously shown by other studies.

Acknowledgements

We especially thank the blood donors for participating in the present study. We express our appreciation to David Goldstein for providing information from published sequences, Meherzia Ben Fadhel and Mònica Vallés for technical support. This work was supported by le Ministère de l'Enseignement Supérieur, de la Recherche Scientifique et de la Technologie de la Tunisie, by Dirección General de Investigación, Ministerio de Ciencia y Tecnología in Spain grant BOS2001-0794, and by Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya grant 2001SGR00285 and fellowship 2000FI00696 to S.P.

References

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J., Staden, R. & Young, I. G. (1981) Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–465.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M. & Howell, N. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147.

Bandelt, H. J., Forster, P., Sykes, B. C. & Richards, M. B. (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743–753.

Bendall, K. E. & Sykes, B. C. (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet* **57**, 248–256.

Bosch, E., Calafell, F., Perez-Lezaun, A., Comas, D., Mateu, E. & Bertranpetit, J. (1997) Population history of North Africa: evidence from classical genetic markers. *Hum Biol* **69**, 295–311.

Bosch, E., Calafell, F., Perez-Lezaun, A., Clarimon, J., Comas, D., Mateu, E., Martinez-Arias, R., Morera, B., Brakez, Z., Akhayat, O., Sefiani, A., Hariti, G., Cambon-Thomsen, A. & Bertranpetit, J. (2000) Genetic structure of north-west Africa revealed by STR analysis. *Eur J Hum Genet* **8**, 360–366.

Bosch, E., Calafell, F., Comas, D., Oefner, P. J., Underhill, P. A. & Bertranpetit, J. (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am J Hum Genet* **68**, 1019–1029.

Brakez, Z., Bosch, E., Izaabel, H., Akhayat, O., Comas, D., Bertranpetit, J. & Calafell, F. (2001) Human mitochondrial DNA sequence variation in the Moroccan population of the Souss area. *Ann Hum Biol* **28**, 295–307.

Brett, M. & Fentress, E. (1996) *The Berbers*. Blackwell, Oxford.

Chaabani, H., Helal, A. N., van Loghem, E., Langaney, A., Benammar Elgaaied, A., Rivat Peran, L. & Lefranc, G. (1984) Genetic study of Tunisian Berbers. I. Gm, Am and Km immunoglobulin allotypes and ABO blood groups. *J Immunogenet* **11**, 107–113.

Chaabani, H. & Cox, D. W. (1988) Genetic characterization and origin of the Tunisian Berbers. *Hum Her* **38**, 308–316.

Close, A. E. & Wendorf, F. (1990) *North Africa at 18000 BP. Vol 2: Low latitudes*. Unwim Hyman, London.

Comas, D., Calafell, F., Benchemsi, N., Helal, A., Lefranc, G., Stoneking, M., Batzer, M. A., Bertranpetit, J. & Sajantila, A. (2000) Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* **107**, 312–319.

Comas, D., Plaza, S., Wells, R. J., Yuldaseeva, N., Lao, O., Calafell, F., Betranpetit, J. (2004) Admixture, migrations and dispersals in Central Asia: Evidence from maternal DNA lineages. *Eur J Hum Genet*.doi: 10.1038/sj.ejhg.5201160.

Côrte-Real, H. B., Macaulay, V. A., Richards, M. B., Hariti, G., Issad, M. S., Cambon-Thomsen, A., Papiha, S., Bertranpetit, J. & Sykes, B. C. (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* **60**, 331–350.

Dupanloup, I. (1993). *Effets de l'adoption de l'agriculture et de l'élevage sur les populations proche-orientales et africaines*. Thesis, University of Geneva.

Efron, B. (1982) *The Jackknife, the bootstrap and other resampling plans*. Philadelphia, Pa: Society for industrial and Applied Mathematics.

Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Ferembach, D. (1985) On the origin of the Iberomaurusians (Upper Palaeolithic: North Africa). A new hypothesis. *J Hum Evol* **14**, 393–397.

Helal, A. N., Rivat-Peran, L., Van Loghem, E., de Lange, G., Langaney, A. & Lefranc, G. (1988) Further contribution of common Gm*-Am* haplotypes and Km* alleles in the characterization of the Tunisian population. *Exp Clin Immunogenet* **5**, 1–14.

Julien, C. A. (1961) Histoire de l'Afrique du Nord. Paris: Payot.

Krings, M., Salem, A. E., Bauer, K., Geisert, H., Malek, A. K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., Sajantila, A., Paabo, S. & Stoneking, M. (1999) MtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? *Am J Hum Genet* **64**, 1166–1176.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B. & Torroni, A. (1999) The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**, 232–249.

Nei, M. (1987) *Molecular evolutionary genetics*. New York: Columbia University Press.

Newman, J. L. (1995) *The peopling of Africa: a geographic interpretation*. New Haven: Yale University Press.

Plaza, S., Calafell, F., Helal, A., Bouzerna, N., Lefranc, G., Bertranpetit, J. & Comas, D. (2003) Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the Western Mediterranean. *Ann Hum Genet* **67**, 312–328.

Quintana-Murci, L., Semino, O., Bandelt, H. J., Passarino, G., McElreavey, K. & Santachiara-Benerecetti, A. S. (1999) Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet* **23**, 437–441.

Rando, J. C., Pinto, F., Gonzalez, A. M., Hernandez, M., Larruga, J. M., Cabrera, V. M. & Bandelt, H. J. (1998) Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern, and sub-Saharan populations. *Ann Hum Genet* **62**, 531–550.

Rando, J. C., Cabrera, V. M., Larruga, J. M., Hernandez, M., Gonzalez, A. M., Pinto, F. & Bandelt, H. J. (1999) Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Ann Hum Genet* **63**, 413–428.

Renfrew, C. (1991) Before Babel: speculations on the origins of linguistic diversity. *Cambrideg Aecheological Journal* **1**, 3–23.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C. & Stefanescu, G. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276.

Richards, M., Macaulay, V., Torroni, A. & Bandelt, H. J. (2002) In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet* **71**, 1168–1174.

Saillard, J., Magalhaes, P. J., Schwartz, M., Rosenberg, T. & Norby, S. (2000) Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. *Hum Biol* **72**, 1065–1068.

Salas, A., Richards, M., De la Fe, T., Lareu, M. V., Sobrino, B., Sanchez-Diz, P., Macaulay, V. & Carracedo, A. (2002). The making of the African mtDNA landscape. *Am J Hum Genet* **71**, 1082–1111.

Thomas, M. G., Weale, M. E., Jones, A. L., Richards, M., Smith, A., Redhead, N., Torroni, A., Scozzari, R., Gratrix, F., Tarekegn, A., Wilson, J. F., Capelli, C., Bradman, N. & Goldstein, D. B. (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* **70**, 1411–1420.

Vigilant, L., Pennington, R., Harpending, H., Kocher, T. D. & Wilson, A. C. (1989) Mitochondrial DNA sequences in single hairs from a southern African population. *Proc Natl Acad Sci U S A* **86**, 9350–9354.

Watson, E. (1996) *Threads from the Past: a genetic study of African ethnic groups and human origins*. PhD dissertation. Massey University, New Zealand.

Watson, E., Forster, P., Richards, M. & Bandelt, H. J. (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* **61**, 691–704.

Received: 26 June 2003
Accepted: 17 December 2003